AAPT PHYSICS EDUCATION | AMERICAN JOURNAL of PHYSICS

Using an online homework system enhances students' learning of physics concepts in an introductory physics course

K. Kelvin Cheng, Beth Ann Thacker, Richard L. Cardenas, and Catherine Crouch

Citation: American Journal of Physics **72**, 1447 (2004); doi: 10.1119/1.1768555
View online: http://dx.doi.org/10.1119/1.1768555
View Table of Contents: http://scitation.aip.org/content/aapt/journal/ajp/72/11?ver=pdfcov
Published by the American Association of Physics Teachers

Using an online homework system enhances students' learning of physics concepts in an introductory physics course

K. Kelvin Cheng,[a] Beth Ann Thacker, and Richard L. Cardenas[b]
Department of Physics, Texas Tech University, Lubbock, Texas 79409-1051

Catherine Crouch[c]
Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138

(Received 27 June 2002; accepted 14 May 2004)

We report the results of a comparison of student understanding of physics concepts with and without online homework, as measured by the force concept inventory. We compared students in large introductory courses taught by interactive engagement and noninteractive engagement methods and with ungraded homework and with online homework. We also compared the understanding of students in different grade subgroups. The increase in the average force concept inventory normalized gain was statistically significant for all students taught with online homework, indicating that graded homework increases student understanding of physics concepts. The gain was significantly higher for those students taught with interactive engagement methods together with online homework. The C grade subgroup taught by interactive engagement methods benefited more from the implementation of online homework than the other subgroups.
[DOI: 10.1119/1.1768555]

I. INTRODUCTION

The decline in the number of physics graduate students at many universities has resulted in a shortage of teaching assistants (TAs) to grade homework or to help students in large multisection introductory physics courses.[1,2] This decline is the case at Texas Tech University (TTU), where most of the TAs teach the introductory physics laboratory classes, which are separate from the lecture courses. The solution to the TA shortage has been either not to grade the homework or to grade only a very small percentage of selected problems, depending on the instructor. Weekly in-class quizzes with problems similar to the assigned, but not graded homework problems, have been given instead.

It often is assumed that doing homework will help students understand physics concepts and that students will not do homework unless it is graded. Because the number of TAs is not expected to increase significantly in the near future, we sought an alternative to the conventional homework system and decided to use an online graded homework system as an alternative and to evaluate the difference in students' understanding of physics concepts with and without its use.

The goals of our study of the use of online graded homework can be summarized as follows:

(1) Determine if the implementation of online homework (OHW) in introductory physics classes leads to an improvement, compared to ungraded homework (UHW), in students' understanding of physics concepts, as measured by the force concept inventory (FCI).[3,4]
(2) Determine if the use of interactive engagement (IE) teaching methods makes a difference in students' understanding of physics concepts, compared to the use of noninteractive engagement (NIE) teaching methods, in the presence or absence of an OHW system.
(3) Determine if a particular letter grade subgroup (A,B, C,D,F) benefited more from an OHW system.

To achieve these goals, the FCI was given as a pre/posttest to students who were taught with and without IE methods, with assigned, but UHW in 1999 and with an OHW system in 2000. Our results based on a two-tail t test of unequal variances indicate that the use of graded OHW significantly enhances student understanding of physics concepts compared to students with UHW, especially in classes taught by IE teaching methods. In addition, the C students taught by IE benefited the most in comparison to the other letter grade subgroups.

II. METHODS AND RESOURCES

A. Student population profile

The course used in this two-year study at TTU was Physics 1308, a multisection calculus-based introductory physics course. Physics 1308 is taught by three to four instructors; one of the faculty members maintains uniformity among all the sections, including a common final exam. The course covers kinematics, force, momentum, energy, and waves. The typical class is ≈250 students, with sections ranging from 35–80 students. The course runs for 14 weeks during the academic year and consists of 2.5 h/week of lecture. There are no recitation sessions, and the laboratory course is separate from the lecture course. The course also is taught in the 21 day summer session, with a lecture of 1.83 h/day. A total of 16 sections from 1999 and 2000 were involved in this study. Some instructors taught more than one section in this study. In 1999, the course grade was calculated from the scores of the exams (80%) and weekly in-class quizzes (20%). In 2000, the OHW counted 10%, in-class quizzes 10%, and the exams counted for 80%. The questions in the exams were divided into conceptual (20%–30%) and problem solving (70%–80%) in all semesters. The cutoff for the grades A, B, and C was 85, 74, and 60, respectively.

The majority of the students in the course are engineering majors, with some students majoring in the physical sciences

Table I. Classification of students into four groups based on teaching methods, IE and NIE, and the use of UHW and an OHW. The number of instructors and the year taught in each group are also indicated.

Group	Teaching methods	Homework	Instructors (year)
UHW/NIE ($N=136$)	UHW	NIE	3 (1999)
UHW/IE ($N=76$)	UHW	IE	2 (1999)
OHW/NIE ($N=86$)	OHW	NIE	4 (2000)
OHW/IE ($N=121$)	OHW	IE	1 (2000)

and in the life sciences. Most students had at least a semester of calculus, a prerequisite of the course, and a high school physics course.

B. Teaching pedagogy and selection of NIE and IE groups

The teaching method was labeled interactive if the instructor attempted to involve the students in thinking about a physics concept during the lecture. Usually, this involvement was done by posing a question to the students, allowing them to discuss it with each other, and then polling their responses by a show of hands.[5] A method was labeled NIE if the instructor used a traditional lecture method with minimal, if any, student participation.

To examine the effects of teaching methods on the students' learning of physics concepts, the answers of students on the pre/post-FCI tests were categorized into four groups as shown in Table I.

C. Internet technology

Online pre/post-FCI tests were incorporated in year 1999, with continued use in 2000; the online web-based course tool (WEBCT)[6] and OHW were implemented in 2000.

During the academic year, all the registered students took online pre/post-FCI tests administered by the Harvard Physics Education Research Group. The FCI pretest was administered during the first week of class. The FCI posttest was administered in the middle of the term, after kinematics, dynamics, and momentum were covered. The FCI normalized gain (g) of each student is the ratio of the absolute gain to the maximum possible gain multiplied by 100:[7]

$$g=\frac{S_{\text{post}}-S_{\text{pre}}}{100-S_{\text{pre}}}\times 100, \tag{1}$$

where S_{pre} and S_{post} are the matched pre- and posttest scores normalized to 100 for each student. The average FCI pretest, $\langle S_{\text{pre}}\rangle$, average posttest, $\langle S_{\text{post}}\rangle$, and average normalized gain. $\langle g\rangle$, and their t statistics[8] were calculated for different groups. The effect size (ES), a measure of the magnitude of the treatment effect between two groups, was calculated from the difference of the means of the two comparison groups, $\mu_1-\mu_2$, divided by the pooled standard deviation. σ_{pooled}. The latter can be calculated from the standard deviations of the two comparison groups, σ_1 and σ_2, as[8]

$$\sigma_{\text{pooled}}=\sqrt{\frac{(n_1-1)\sigma_1^2+(n_2-1)\sigma_2^2}{n_1+n_2-2}}, \tag{2}$$

where n_1 and n_2 are the sample sizes of the two comparison groups.

An in-class written version of the FCI was given to the students in summer 2000 and fall 2000. Both internal and external studies demonstrate that there is no appreciable difference in the FCI scores based on the type of administration.[9] We report results only of the students who participated in both the pre- and posttest. The FCI scores from the 20%–30% of the students who dropped the course before the final exam or missed the FCI pre- or posttest are not reported. A total of 212 students and 207 students participated in both the pre- and post FCI tests in 1999 and 2000, respectively.

In 2000, we used the online course management tool, WEBCT. The features of WEBCT that we used most frequently were the online grade book module for the posting of OHW, in-class quiz and exam scores, the web chat module for interactive real-time communication with students, and the online quiz/survey module for posting practice quizzes, surveys, and questions. The web chat feature was used almost daily in the summer 2000 section taught using IE. It was used much less frequently in the other sections taught by either method. WEBCT also was used as a centralized internet resource with hyperlinks for the students to access the FCI, OHW, and other resources, such as lecture notes, worksheets, and keys to quizzes and sample problems.

In 2000, we used an OHW system provided by the University of Texas at Austin. This system allows instructors to select questions, numeric or multiple choice, from a databank of several thousand physics questions. Most of these questions have multiple parts and the numerical variables in the questions are generated randomly to ensure that each student obtains a different version of the same question with an unique solution. About 10–15 online questions were assigned to the students each week with about 10%–20% of the questions being conceptual in nature. Students had to log on to the OHW server to download the questions. The system provides instant feedback for right or wrong answers, and students have between five and ten trials to obtain the correct answer. Each time a student obtains a wrong answer, the maximum possible score is decreased. The OHW constituted 10% of their total grade. The year 2000 FCI statistics were only from students who participated in the OHW.

D. Other resources

Voluntary help sessions sponsored by the engineering school and the physics department were available in 1999 and 2000. The instructor for the help session sponsored by the engineering school was an engineering undergraduate senior who conducted two regular 2 h sessions per week for each semester in 1999 and 2000. The instructor for the other help session was a physics graduate student who conducted a one 2 h session per week in Spring 1999; it was discontinued due to the lack of student participation and financial resources. However, the physics department provided centralized TA office hours (4–5 h/day during the week) in 1999 and 2000. No particular tutorial training or structured instructions were given to the undergraduate or graduate students in this study. Participation in the help sessions was very poor, with no more than ten students showing up for the regular 2 h sessions in 1999, except the sessions before the exams. Only a handful of students showed up regularly at the centralized TA office. However, there was a significant in-

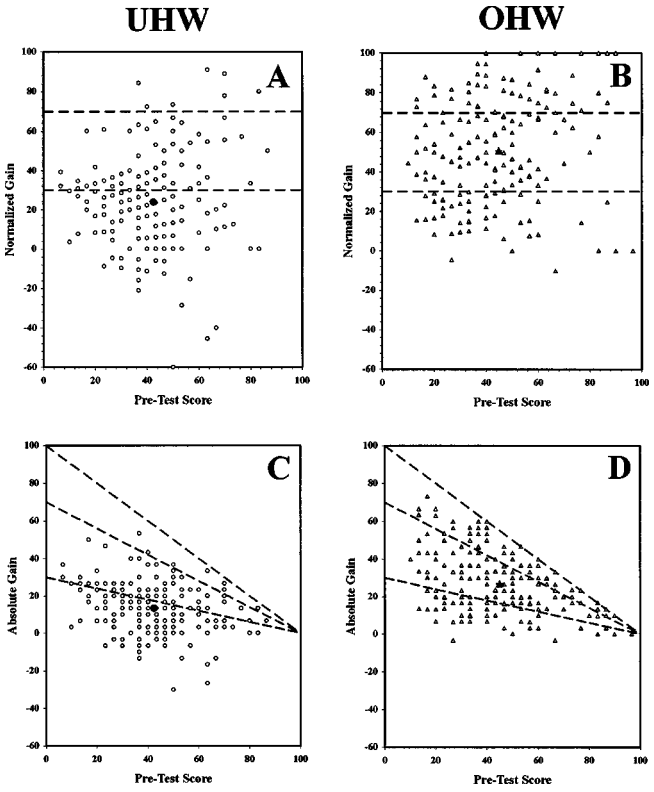

Fig. 1. FCI data of TTU students with UHW and OHW. Plots of FCI normalized gain (g) versus pretest score (S_{pre}) of TTU students with UHW (open circle) (A) and with OHW (open triangle) (B) are shown. The two horizontal dashed lines indicate the cutoff between low-to-medium normalized gain at 30% and medium-to-high normalized gain at 70% (according to Ref. 7). Plots of FCI absolute gain ($S_{\text{post}}-S_{\text{pre}}$) versus pretest score ($S_{\text{pre}}$) with UHW (open circle) (C) and with OHW (open triangle) (D) also are shown. The upper, middle, and lower dashed slopes indicate the 100%, 70%, and 30% normalized gain boundaries. The means of students taught by UHW and OHW are given by solid circle and solid triangle, respectively.

crease in the number of students attending the help sessions, ≈30–50, as well as those seeking help at the centralized TA office, in 2000, because homework was graded and constituted 10% of their total grade. In the Summer 2000 section, these resources were not available. The instructor offered online office hours in addition to regular office hours, in lieu of centralized TA office hours and help sessions.

III. RESULTS

A. Comparisons of FCI scores: All data

The distributions of the FCI scores, normalized gain, g, versus pretest score, S_{pre}, for students with and without OHW are shown in Figs. 1(a) and 1(b). The average normalized gain, $\langle g\rangle$, of the OHW group was between 30% and 70%, which is within the low-to-medium gain region.[7] On the other hand, $\langle g\rangle$ of the UHW group was <30%, which is in the low gain region.[7] The absolute gain versus the pretest score for students with UHW and graded homework is shown in Figs. 1(c) and 1(d). The distributions of the FCI scores for different grade subgroups are shown in Fig. 2. The means and standard errors (SEs) of the FCI scores also are shown. The average normalized gain of the grade subgroups was in the low gain region except the A grade subgroup in the UHW group. In contrast, $\langle g\rangle$ for all grade subgroups in the OHW group was within the low-to-medium gain range.

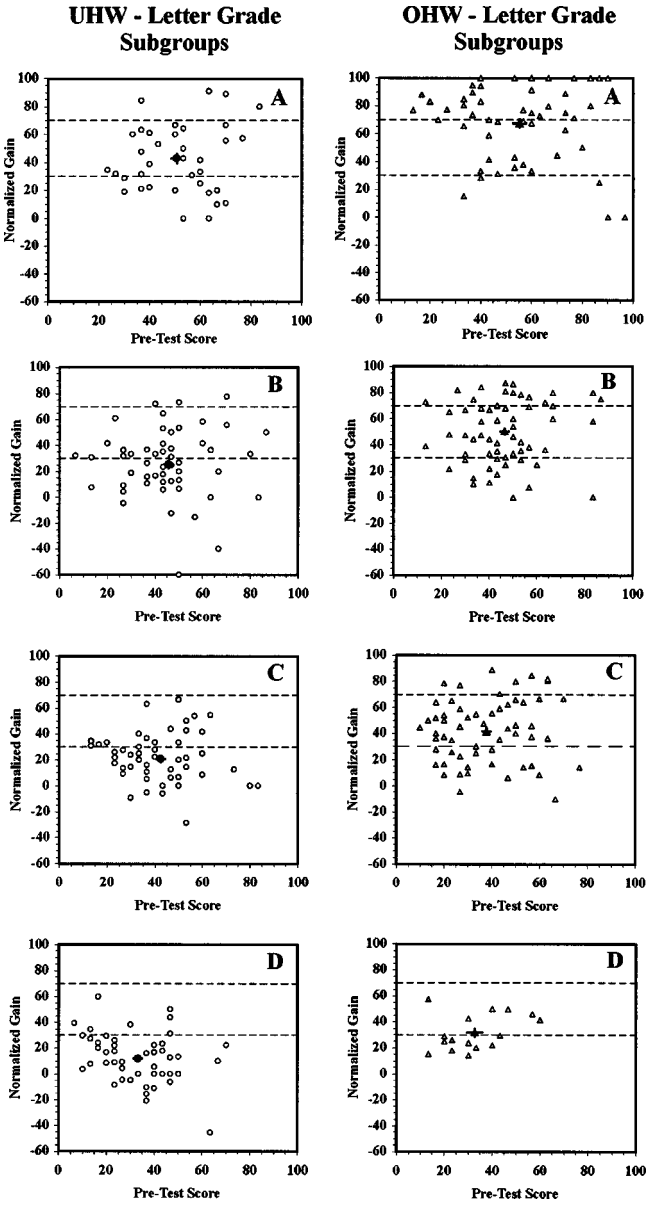

Fig. 2. FCI data of TTU students from different letter grade subgroups with UHW and OHW. Plots of FCI normalized gain versus pretest score of TTU students from different letter grades (A, B, C, and D) subgroups with UHW (open circle) and OHW (open triangle) are shown. See the caption of Fig. 1 for other details.

The average FCI results (means±SE) for all students in the UHW group ($N=212$) were $\langle S_{\text{pre}}\rangle=42.4\pm1.1$, $\langle S_{\text{post}}\rangle=55.8\pm1.3$, and $\langle g\rangle=23.9\pm1.7$, respectively, compared with $\langle S_{\text{pre}}\rangle=44.9\pm1.3$, $\langle S_{\text{post}}\rangle=71.8\pm1.3$, and $\langle g\rangle=50.9\pm1.8$, respectively, for all students in the OHW group ($N=207$). Based on the two-tailed t test with unequal variances,[8] the difference between the UHW and the graded homework average pretest scores is statistically insignificant; the probability that the means of the two populations are the same within statistical variation, $P(t\leqslant t^*)=0.158$, is greater than 0.05 corresponding to a 95% confidence threshold level, where t^* is the calculated t value from the two comparing populations. The difference between the ungraded and graded average posttest scores is statistically significant, with a vanishing small probability that the means of the two popu-

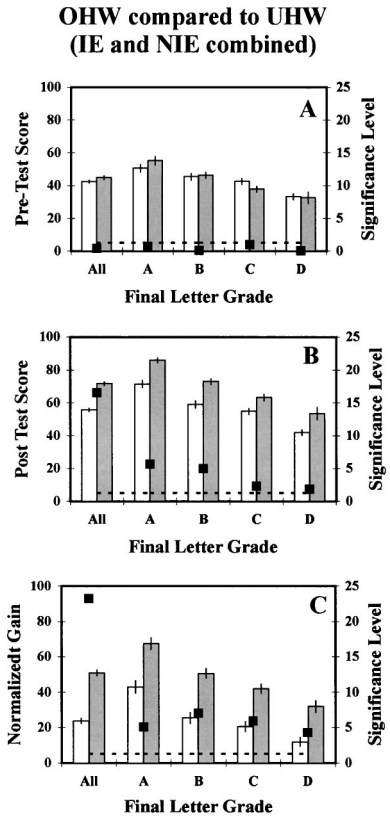

Fig. 3. Comparisons of FCI statistics of TTU students with UHW and with OHW. Average FCI pretest (Panel A), posttest (Panel B), and normalized gain (Panel C) scores of students with UHW (white bars) and students with OHW (shaded bars) are shown. Averages are shown for all students and for groups of students earning final letter grades of A, B, C, and D or lower (letter grade subgroups), as indicated below the horizontal axes. Vertical lines on the top of the histogram bars represent the standard errors. The *t*-test significance level (solid square) and 95% significance threshold ($\alpha=0.05$; dotted line) also are shown. The significance level is defined as $-\log(P(t \leqslant t^*))$ as discussed in the text.

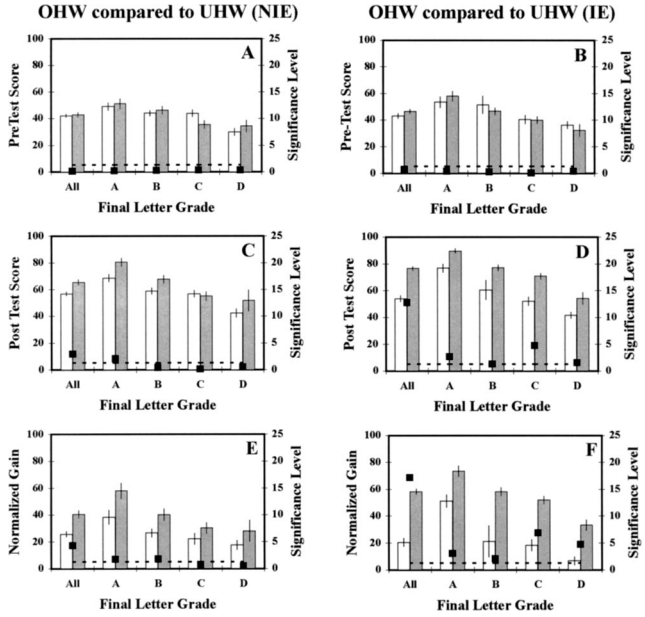

Fig. 4. Comparisons of FCI statistics of TTU students taught by NIE methods with those by IE methods with UHW and OHW. Average FCI pretest (Panels A and B), posttest (Panels C and D), and normalized gain (Panels E and F) scores of students with UHW (white bars) and students with OHW (shaded bars) for those taught by NIE methods (Panels A, C, and E) and by IE methods (Panels B, D, and F) are shown. Averages are shown for all students and for groups of students earning letter grades of A, B, C, and D or lower, as indicated below the horizontal axes. Vertical lines on the top of the histogram bars represent the standard errors. The *t* test significance level (solid square) and 95% significance threshold ($\alpha=0.05$; dotted line) are shown. The significance level is defined as $-\log(P(t \leqslant t^*))$.

lations are the same, $P(t \leqslant t^*) = 2.9 \times 10^{-17}$. The difference between the average normalized gains is also statistically significant, with $P(t \leqslant t^*) = 6.3 \times 10^{-24}$.

To facilitate a comparison of the FCI scores of different groups, we define the significance level parameter as $-\log_{10}(P(t \leqslant t^*))$. The significance level threshold of 0.05 corresponding to a 95% confidence level is 1.3. The significance level parameter increases with an increasing level of statistical significance between two population means. The difference between two population averages is significant only if the significance level is greater than the threshold of 1.3. The levels of statistical significance of the differences between the averages of any two comparable sample populations can be plotted directly as shown in Figs. 3–5. For example, the significance level parameters between the FCI score averages for the UHW and graded homework group were 0.81 (<1.3), 16.5 (>1.3), and 23.2 (>1.3), for the pretest, posttest, and normalized gain, respectively (see Fig. 3).

Another common measure of statistical significance is the ES (see Eq. (2)). A positive ES indicates an improvement in the first group relative to the second, for example, a positive ES indicates higher performance of the OHW group. An ES less than 0.5 indicates a small effect, between 0.5 and 0.8, a medium effect, and greater than 0.8, a large effect. The effective sizes for the FCI data of different subgroups are shown in Table II.

B. Comparisons of FCI scores for letter grade subgroups

The question of whether the observed increase in the FCI scores, which was found with the OHW students, differed for students in different grade subgroups was examined. Table III gives the means and SE of the FCI, posttest and normalized gain for all students as well as for students divided into groups according to their final grades. The pretest scores in the UHW and graded homework groups were the same within statistical significance for each subgroup. However, the posttest and normalized gain scores were higher for the OHW group for the student population as a whole and for each grade subgroup, as shown in Fig. 3 and Table III. In addition, the percentage of A, B, and C students rose dramatically when UHW was replaced by online graded homework and the percentage of D students decreased, while the difficulty of the course exams was kept constant.

We determined the significance level parameter for the difference between the FCI average scores with UHW and graded homework for each grade subgroup. For all grade subgroups, the significance level parameters of the pretest were all below the threshold value as shown in Fig. 3(a). For the posttest and normalized gain, the significance level parameters were all above the threshold [see Figs. 3(b) and 3(c)]. As shown in Table II, the ESs of $\langle g \rangle$ of all grade subgroups were all greater than 0.8, indicating a very substantial impact of OHW on the improvement of FCI gain.

Table II. ESs for comparisons of FCI data. ESs for comparison FCI data of groups or subgroups with OHW to those with UHW and also the data of NIE subgroups with that of IE subgroups. ESs greater than 0.8 are indicated by asterisks.

	(a) OHW compared to UHW		
	$\langle S_{pre} \rangle$	$\langle S_{post} \rangle$	$\langle g \rangle$
All	0.14	0.86*	1.05*
A	0.25	1.09*	0.97*
B	0.05	0.82*	1.01*
C	−0.29	0.51	0.93*
D	−0.04	0.79	1.14*
NIE	0.04	0.47	0.59
IE	0.18	1.25*	1.51*
NIE—A	0.14	1.00*	0.88*
NIE—B	0.18	0.65	0.73
NIE—C	−0.62	−0.11	0.45
NIE—D	0.38	0.72	0.75
IE—A	0.27	1.77*	1.37*
IE—B	−0.33	1.16*	1.76*
IE—C	−0.03	1.68*	2.33*
IE—D	−0.34	1.14*	2.17*
	(b) IE compared to NIE		
UHW	0.06	−0.13	−0.20
OHW	0.18	0.62	0.72
UHW—A	0.28	0.59	0.55
UHW—B	0.43	0.10	−0.20
UHW—C	−0.24	−0.31	−0.17
UHW—D	0.44	−0.05	−0.61
OHW—A	0.33	0.78	0.62
OHW—B	0.03	0.63	0.82*
OHW—C	0.27	0.97*	1.05*
OHW—D	−0.13	0.15	0.38

C. Comparisons of FCI scores for the NIE and IE groups

We distinguished the effects of the homework system and pedagogy by comparing FCI scores for those with UHW to those with OHW for those students taught only by NIE or those taught only by IE methods. Although there was no significant difference between the pretest scores of students in these various groups, the posttest and normalized gain with OHW were higher than with UHW for both teaching methods, as shown in Table III. However, the significance level parameters of the posttest [Fig. 4(d)] and the normalized gain [Fig. 4(f)] comparing UHW to an OHW system were much higher for students in the IE group than for students in the NIE group [Figs. 4(c) and 4(e)]. As shown in Table II, the ESs of $\langle S_{post} \rangle$ and $\langle g \rangle$ for IE were both greater than 0.8, whereas those for NIE were less than 0.8. These results agree with the higher significant level parameters of all letter grades for IE [Figs. 4(d) and 4(f)] than those for NIE group [Figs. 4(c) and 4(e)].

Among grade subgroups in the NIE group [Figs. 4(c) and 4(e)], the significance level parameters of the posttest for the A subgroup and the normalized gain of the A and B subgroups were only slightly higher than the threshold value. In contrast, for all grade subgroups in the IE group [Figs. 4(d) and 4(f)], all significance level parameters of the posttest and normalized gain were higher than the threshold. The significance level parameters of posttest and normalized gain for the C subgroup in the IE group were several times higher than those for the other subgroups within the IE group and all subgroups in the NIE group. Again, the results of ESs supported the *t* statistics. As shown in Table II, the ESs of $\langle S_{post} \rangle$ and $\langle g \rangle$ for the IE grade subgroups were all greater than 0.8, while those for the NIE subgroups were less than 0.8. In addition, the ES of $\langle S_{post} \rangle$ or $\langle g \rangle$, 1.68 or 2.33, respectively, for the IE-C subgroup was the highest among the IE and NIE grade subgroups as shown in Table II.

D. Comparisons of FCI scores in NIE groups with those in IE groups

The comparison of the FCI scores from students taught with IE methods to those taught with NIE methods in either ungraded or graded groups provided further information that distinguishes the effect of homework from the effect of pedagogy. Again, there was no significant difference in the pretest scores of students taught by IE and NIE methods in either year, as shown in Figs. 5(a) and 5(b). For students in the UHW group, the significance level parameters comparing IE students and NIE students for all students and different grade level subgroups were all below or near the threshold, indicating no statistically significant difference between the IE and NIE scores [see Figs. 5(c) and 5(e)]. In contrast, for students in the OHW group, the significance level parameters of the posttest and the normalized gain for all students and the A, B, and C grade subgroups were at or above threshold as shown in Figs. 5(d) and 5(f). In particular, the scores of the C subgroup were significantly higher with the OHW system. The results of ESs also supported the *t* statistics. As shown in Table II, the effect size of $\langle S_{post} \rangle$ and $\langle g \rangle$, 0.97 and 1.05, respectively, in the OHW C subgroup was the highest among the UHW and the OHW grade subgroups.

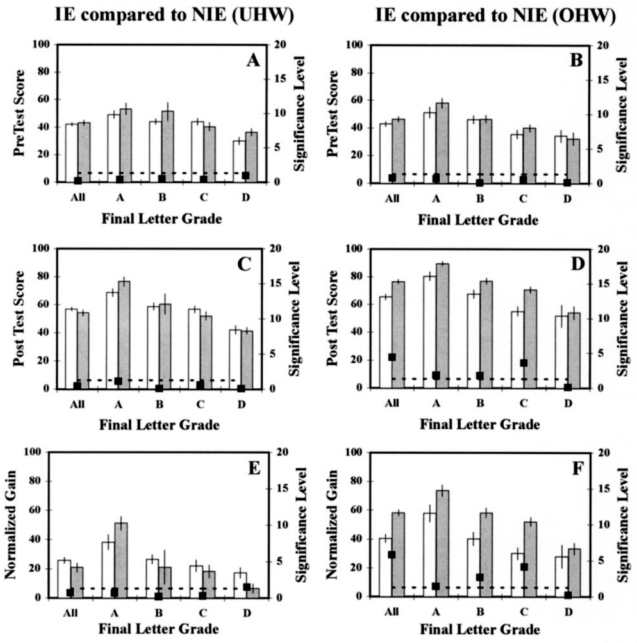

Fig. 5. Comparisons of FCI statistics of TTU students taught by UHW with those with OHW by NIE and IE methods. Average FCI pretest (Panels A and B), posttest (Panels C and D), and normalized gain (Panels E and F) scores of students taught with NIE methods (white bars) and students taught with IE methods (shaded bars) with UHW (Panels A, C, and E) and OHW (Panels B, D, and F) are shown. Averages are shown for all students and for groups of students earning grades of A, B, C, and D or lower. Vertical lines on the top of the histogram bars represent the standard errors. The *t* test significance level (solid square) and 95% significance threshold (α=0.05; dotted line) are shown. The significance level is defined as $-\log(P(t \leq t^*))$.

Table III. Summary of FCI pretest, posttest, and normalized gain (means±SE) scores with UHW and graded homework. The means and SEs of FCI pretest (first row), posttest (second row), and normalized gain (third row) of students with UHW and OHW as a whole (all), in IE and NIE groups, and in letter grade subgroups are shown. The population size (N) of each group or subgroup is given and the percentage of the total population in each subgroup. The ratings of the normalized gain scores, that is, low gain (0–30), medium gain (30–70), and high gain (70–100), based on the recently published extensive FCI survey by Hake are indicated by +, ++, and +++, respectively.[a] The FCI scores of the Summer 2000 section (^{s}), a subset of the IE and OHW group also are shown.

Homework	Teaching methods	All letter grades	A subgroup	B subgroup	C subgroup	D subgroup
UHW	All	42.3±1.1	50.6±2.4	45.5±2.2	42.5±2.0	33.2±2.0
		55.8±1.3	71.4±2.3	59.2±2.4	54.9±2.0	41.9±2.0
		23.9±1.7	43.0±3.8	25.5±3.4	20.6±3.0	11.9±2.6
		$N=212$	$N=39$ (18%)	$N=61$ (29%)	$N=57$ (27%)	$N=55$ (26%)
		+	++	+	+	+
	NIE	42.1±1.4	49.1±2.9	44.0±2.2	43.9±2.6	30.0±2.7
		56.8±1.5	68.6±1.5	58.8±2.9	56.7±2.4	42.3±3.2
		25.7±2.0	38.3±5.2	26.5±3.2	22.0±4.0	17.4±3.6
		$N=136$	$N=25$ (18%)	$N=49$ (36%)	$N=35$ (26%)	$N=27$ (20%)
		+	++	+	+	+
	IE	43.1±2.0	53.3±4.3	51.4±6.8	40.3±3.3	36.3±2.7
		54.0±2.3	76.7±3.3	60.6±7.3	52.0±3.2	41.5±2.4
		20.5±3.2	51.2±4.7	21.1±11.6	18.2±4.4	6.67±3.2
		$N=76$	$N=14$ (18%)	$N=12$ (16%)	$N=22$ (29%)	$N=28$ (37%)
		+	++	+	+	+
OHW	All	44.9±1.3	55.2±2.8	46.4±2.0	37.8±2.0	32.7±3.5
		71.8±1.3	86.0±1.7	73.0±1.9	63.4±2.2	53.6±3.9
		50.9±1.8	67.5±3.5	50.6±2.9	42.0±2.8	32.0±3.5
		$N=207$	$N=56$ (27%)	$N=67$ (32%)	$N=68$ (33%)	$N=16$ (7%)
		++	++	++	++	++
	NIE	42.9±1.9	51.1±3.9	46.1±3.1	35.5±3.1	34.2±4.6
		65.5±2.1	80.4±3.1	67.6±3.0	55.1±3.3	51.8±8.0
		40.4±2.0	58.0±6.0	40.1±4.5	30.0±4.2	27.9±8.1
		$N=86$	$N=22$ (23%)	$N=28$ (33%)	$N=32$ (37%)	$N=4$ (5%)
		++	++	++	++	+
	IE	46.4±1.8	57.9±3.8	46.6±2.6	39.9±2.6	32.2±4.6
		76.4±1.5	89.6±1.6	76.9±2.2	70.7±2.2	54.2±4.6
		58.2±2.1	73.7±4.1	58.2±3.3	52.0±3.0	33.3±3.9
		$N=121$	$N=34$ (28%)	$N=39$ (32%)	$N=36$ (30%)	$N=12$ (10%)
		++	+++	++	++	++
		40.2 ± 3.5^s	53.4 ± 12.1^s	43.8 ± 4.2^s	32.1 ± 4.2^s	18.3 ± 5.0^s
		75.5 ± 2.4^s	91.7 ± 2.1^s	85.5 ± 2.3^s	64.2 ± 3.3^s	53.3 ± 10.0^s
		59.9 ± 3.4^s	80.7 ± 1.9^s	64.0 ± 4.4^s	46.6 ± 4.6^s	41.9 ± 15.8^s
		$N=33^s$	$N=6^s$ (18%)	$N=14^s$ (42%)	$N=11^s$ (33%)	$N=2^s$ (6%)

[a]Please see Ref. 7.

E. Two-factor ANOVA analysis of the combined effects of teaching methods and homework on FCI scores

A two-factor ANOVA analysis allows us to investigate the effects of two independent factors or treatments, teaching methods (IE versus NIE) and homework (OHW versus UHW) on a dependent variable. Also, the interaction effect of the various treatments can be examined directly. We have chosen the pretest or normalized gain as the dependent variable and have generated the two-factor ANOVA as shown in Table IV. For the pretest as the dependent variable, the P values for the effects of teaching method, homework, and their interaction (the product of the teaching method and homework) were all greater than 0.05 (or 95% confidence interval), indicating the lack of statistical significance of the pretest scores with the two treatments and their interaction. On the other hand, with the normalized gain as the dependent variable, the P values for the two treatments and their interaction were significantly less than 0.05, indicating the existence of strong and statistically significant effects of the individual treatment and the interaction of the two treatments on the students' understanding of physics concepts as given by the normalized gain scores.

IV. DISCUSSION

Although it is widely assumed that doing homework is an important aspect of learning physics, we know of little literature on the effect of homework on students' learning of physics concepts at the university level.[10–15] Studies that have compared the understanding of students with OHW to students with paper-and-pencil graded homework found no significant difference[12] based on exam performance.[13] We have compared the understanding of students assigned online graded homework with students assigned UHW. Although we have not addressed the question of whether it is the process of doing the homework (providing students with a motivation to think critically about physics concepts), the feedback of graded homework, or some other aspect of homework, we have found that the results indicate that

Table IV. Two-factor ANOVA analysis of the FCI data. The effects of two teaching methods (NIE versus IE) and homework (OHW versus UHW), and their interaction (the product of the teaching method and homework) on a single dependent variable, pretest or normalized gain, was analyzed using a two-factor ANOVA.[a] The statistical significance of the individual or interaction effects was determined by the P value (<0.05) based on a 95% confidence level.

Dependent variable: Pretest	Sum of squares	Degree of freedom	Mean square	F	P value
Teaching Methods	468.159	1	468.159	1.465	0.227
Homework	418.82	1	418.82	1.311	0.253
Teaching Methods and Homework	173.75	1	173.75	0.544	0.461
Error	132589.821	415	319.494		
Total	932556.09	419			
Dependent variable: Normalized gain	**Sum of squares**	**Degree of freedom**	**Mean square**	F	P value
Teaching Methods	3953.967	1	3953.967	6.388	[b] 1.186×10^{-2}
Homework	67805.839	1	67805.839	109.547	[b] 6.585×10^{-23}
Teaching Methods and Homework	13200.156	1	13200.156	21.326	[b] 5.175×10^{-6}
Error	256870.49	415	618.965		
Total	929362.29	419			

[a]See Ref. 8.
[b]Statistically significant with P-value <0.05, or greater than 95% confidence.

graded homework is a significant factor in increasing students' understanding of physics concepts. In addition, the interactive engagement teaching methods in this study did not increase students' understanding of concepts when they were not coupled with graded homework. The coupling of graded homework (in this case, OHW) and IE teaching methods raised students' understanding much more than the addition of graded OHW to NIE methods, although there was a significant change in the latter case also.

ACKNOWLEDGMENTS

This work was supported by grants from the Robert A. Welch Research Foundation (D-1158) and the Teaching Learning and Technology Center of TTU to one of the authors (K.K.C.). The secure online FCI was developed at Harvard University by CHC and Eric Mazur with support from the National Science Foundation through Award No. DUE-9980802. The valuable consultation and assistance of Dr. Richard P. McGlynn of TTU in constructing the unbalanced two-factor ANOVA tables are gratefully acknowledged.

[a)]Electronic mail: kelvin.cheng@ttu.edu
[b)]Present address: St. Mary's University, Department of Physics, San Antonio, TX 78228.
[c)]Present address: Department of Physics and Astronomy, Swarthmore College, Swarthmore, PA 19081.

[1]AIP Statistics Division, Enrollments and Degrees Report, "First-year U.S. and foreign graduate physics students, 1965 to 1999," American Institute of Physics, ⟨www.aip.org/statistics/trends/highlite/ed/figure5.htm⟩.
[2]At other universities represented by the authors, there are a sufficient number of TA's, but insufficient funds to pay them.
[3]D. Hestenes, M. Wells, and G. Swackhamer, "Force concept inventory," Phys. Teach. **30**, 141–157 (1992).
[4]I. A. Halloun and D. Hestenes, "Common-sense concepts about motion," Am. J. Phys. **53**, 1056–1065 (1985).
[5]E. Mazur, *Peer Instruction* (Prentice–Hall, Upper Saddle River, N.J., 1997), pp. 9–18.
[6]WEBCT, ⟨www.webct.com⟩, is a Web Shell with several built-in telecommunication modules and numerous features for course management used by both the student and the instructor.
[7]R. R. Hake, "Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses," Am. J. Phys. **66**, 64–74 (1998).
[8]George A. Ferguson, *Statistical Analysis in Psychology and Education* (McGraw–Hill, N.Y., 1971), pp. 107–119.
[9]D. MacIsaac, R. P. Cole, D. M. Cole, L. McCullough, and J. Maxka, "Standardized testing in physics via the World Wide Web," *Electronic Journal of Science Education 6(3)* March 2002 (http://unr.edu/homepage/crowther/ejse/ejsev6n3.html).
[10]W. T. Griffith, "Factors affecting performance in introductory physics courses," Am. J. Phys. **53**(9), 839–842 (1985).
[11]M. J. Marr, E. W. Thomas, M. R. Benne, A. Thomas, and R. M. Hume, "Development of instructional systems for teaching and electricity and magnetism course," Am. J. Phys. **67**(9), 789–802 (1999).
[12]S. Bonham and R. Beichner, "Online homework: Does it make a difference?" Phys. Teach. **39**(5), 293–297 (2001).
[13]R. Dufresne, J. Mestre, D. M. Hart, and K. A. Rath, "The effect of web-based homework on test performance in large enrollment introductory physics courses," J. Comput. Math. Sci. Teach. **21**(3), 229–251 (2002).
[14]M. Thoennessen and M. J. Harrison, "Computer-assisted assignments in a large physics class," Comput. Educ. **27**(2), 141–147 (1996).
[15]P. E. Sokol, "Improvements in introductory physics courses. Emphasizing homework and deemphasizing exams can make quantum improvements in learning," New Dir. Instit. Res. **20**(78), 41–44 (1993).